

Ergodicity, transitivity, and regularity for linear cellular automata over \mathbf{Z}_m ¹

Gianpiero Cattaneo^{a,*}, Enrico Formenti^b, Giovanni Manzini^c,
Luciano Margara^d

^a *Università di Milano, Dipart. di Scienze della Informazione, via Comelico 39,
I-20135 Milano, Italy*

^b *Università di Milano, Dipart. di Scienze dell'Informazione, Via Comelico 39, I-20135 Milano, Italy*

^c *Università di Torino, Dipart. di Scienze e Tecnologie Avanzate, Torino, Italy*

^d *University of Bologna, Dipart. di Scienze dell'Informazione, Mura Antez zamboni 7,
40127 Bologna, Italy*

Received September 1997; revised December 1997

Communicated by M. Nivat

Abstract

We study the dynamical behavior of D -dimensional linear cellular automata over \mathbf{Z}_m . We provide an easy-to-check necessary and sufficient condition for a D -dimensional linear cellular automata over \mathbf{Z}_m to be *ergodic and topologically transitive*. As a byproduct, we get that for linear cellular automata ergodicity is equivalent to topological transitivity. Finally, we prove that for 1-dimensional linear cellular automata over \mathbf{Z}_m , regularity (denseness of periodic orbits) is equivalent to surjectivity. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Discrete time dynamical systems; Cellular automata; Ergodicity; Topological transitivity

1. Introduction

Cellular automata (CA) are dynamical systems consisting of a regular lattice of variables which can take a finite number of discrete values. The state of the CA, specified by the values of the variables at a given time, evolves in synchronous discrete time steps according to a given *local rule*. CA can display a rich and complex temporal evolution whose exact determination is in general very hard, if not impossible. In particular, many properties of the temporal evolution of general CA are undecidable [6, 7, 14]. Despite their simplicity that allows a detailed algebraic analysis, linear CA

* Corresponding author.

E-mail address: cattang@dsi.unimi.it (G. Cattaneo)

¹ A preliminary version of this paper has been presented to the Symposium on Theoretical Aspects of Computer Science (STACS'97).

over \mathbf{Z}_m exhibit many of the complex features of general CA. Several important properties of linear CA have been studied during the last few years (see, e.g., [2, 9, 10, 13]) and in some cases exact results have been carried out. As an example, in [13] the authors present criteria for surjectivity and injectivity of the global transition map of linear CA.

The *qualitative* dynamical behavior of CA is a main subject in CA theory. Quoting from [13]:

“Criteria are desired for determining when the sequence of transations of a state configuration of a cellular automata takes a certain type of dynamical behavior.”

In this paper we study the dynamical behavior of linear CA over \mathbf{Z}_m in the framework of ergodic theory. Ergodic theory has been recently applied to CA in a number of works. Some preliminary results can be found in [12, 16–18]. We solve the two following problems: (1) how to decide whether the global transition map of a given D -dimensional linear CA over \mathbf{Z}_m is *ergodic*, and (2) how to *construct* ergodic D -dimensional linear CA over \mathbf{Z}_m . We prove (Theorem 3.2) that a given D -dimensional CA over \mathbf{Z}_m is ergodic if and only if the greatest common divisor of *certain* coefficients of the local rule associated to the CA and m is equal to 1. We wish to emphasize that this result has been independently proved by Sato in [15] and by the authors of this paper in [3]. The proof technique used in [15] relies on a measure theoretic argument and is completely different from our proof which relies on a simple algebraic characterization of ergodic endomorphisms of compact abelian groups given in [16].

This result sheds some light on the role played by selected coefficients as far as topological properties are concerned, emphasizing the different nature of set theoretic properties, such as injectivity and surjectivity, and dynamical properties, such as ergodicity and topological transitivity. The solution of problem (1) generalizes a result presented in [16] (Corollary 2, p. 406), while the solution of problem (2) answers a question raised in [17] (Question 2, p. 605).

We establish a connection between ergodic theory and topological chaos in the case of CA. Although a universally accepted definition of chaos does not exist, two properties are widely accepted as important features of chaotic behavior: topological transitivity and sensitivity to initial conditions. Sensitivity is recognized as a central notion in chaos theory because it captures the feature that in chaotic systems small errors in experimental readings lead to large scale divergence, i.e., the system is unpredictable.

Topological transitivity guarantees that the system cannot be decomposed into two or more subsystems which do not interact under iterations of the map. We prove (Theorem 3.2) that a linear CA over \mathbf{Z}_m is ergodic if and only if it is topologically transitive. Since in [4] one of the author proved that topologically transitive CA are sensitive to initial conditions, we conclude that, for linear CA, ergodicity is equivalent to topological chaos. In Theorem 3.1 we take advantage of the compactness of the space of the configurations on which CA are defined for proving that topologically transitive CA are surjective.

Finally, we study another widely accepted feature of chaotic behavior: *denseness of periodic orbits* (see, e.g., [8]) sometimes referred to as *regularity*. Here (Theorem 3.4) we prove that for 1-dimensional linear CA over \mathbf{Z}_m regularity is equivalent to surjectivity, which is a generalization of the results contained in [9] (Theorem 4) and in [5].

The rest of this paper is organized as follows. In Section 2 we give basic definitions and notations. In Section 3 we list our results. Section 4 contains the proofs of the theorems stated in Section 3. Section 5 contains some indications for further works.

2. Basic definitions

Let \mathbf{Z} and \mathbf{N} denote the set of integers and natural numbers, respectively. Let \mathbf{Z}_m , $m \geq 2$, denote the finite commutative ring of integers modulo m . Throughout the paper we will add and multiply elements of \mathbf{Z}_m by using modular arithmetic. We consider the *space of configurations*

$$\mathbf{Z}_m^{\mathbf{Z}^D} = \{c \mid c: \mathbf{Z}^D \rightarrow \mathbf{Z}_m\}.$$

Each element of $\mathbf{Z}_m^{\mathbf{Z}^D}$ can be visualized as an infinite D -dimensional lattice in which each cell contains an element of \mathbf{Z}_m .

Let $s \geq 1$ and $f: \mathbf{Z}_m^s \rightarrow \mathbf{Z}_m$ be any map. We say that s is the size of the domain of f , or simply the size of f . A *neighborhood frame* of size s is an ordered set of D -dimensional vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_s$. A D -dimensional CA based on a *local rule* f is a pair $(\mathbf{Z}_m^{\mathbf{Z}^D}, F)$, where

$$F: \mathbf{Z}_m^{\mathbf{Z}^D} \rightarrow \mathbf{Z}_m^{\mathbf{Z}^D},$$

is the *global transition map* defined as follows. For every $c \in \mathbf{Z}_m^{\mathbf{Z}^D}$ the configuration $F(c)$ is such that for every $\mathbf{v} \in \mathbf{Z}^D$

$$[F(c)](\mathbf{v}) = f(c(\mathbf{v} + \mathbf{u}_1), \dots, c(\mathbf{v} + \mathbf{u}_s)). \quad (1)$$

In other words, the content of cell \mathbf{v} in the configuration $F(c)$ is a function of the content of cells $\mathbf{v} + \mathbf{u}_1, \dots, \mathbf{v} + \mathbf{u}_s$ in the configuration c . Note that the local rule f and the neighborhood frame completely determine F .

We say that the map $f: \mathbf{Z}_m^s \rightarrow \mathbf{Z}_m$ is linear over \mathbf{Z}_m if and only if there exist $\lambda_1, \dots, \lambda_s \in \mathbf{Z}_m$ such that

$$f(x_1, \dots, x_s) = \sum_{i=1}^s \lambda_i x_i \bmod m.$$

We say that a CA defined over \mathbf{Z}_m is linear if the local rule on which it is based is linear. Note that for a linear D -dimensional CA over \mathbf{Z}_m , Eq. (1) becomes

$$[F(c)](\mathbf{v}) = \sum_{i=1}^s \lambda_i c(\mathbf{v} + \mathbf{u}_i) \bmod m. \quad (2)$$

Example 1. We describe a 2-dimensional linear CA over \mathbf{Z}_2 based on a local rule f which computes the sum modulo 2 of its 4 input. The neighborhood frame of each cell consists of its north, west, east, and south neighbors. Formally, we have $D = 2$, $s = 4$, and

$$\mathbf{u}_1 = (0, 1), \quad \mathbf{u}_2 = (-1, 0), \quad \mathbf{u}_3 = (1, 0), \quad \mathbf{u}_4 = (0, -1).$$

The local rule f is defined by $f(x_1, x_2, x_3, x_4) = (x_1 + x_2 + x_3 + x_4) \bmod 2$. The global transition map F is defined by

$$[F(c)](i, j) = (c(i, j+1) + c(i-1, j) + c(i+1, j) + c(i, j-1)) \bmod 2. \quad \square$$

For linear 1-dimensional CA over \mathbf{Z}_m we use a simplified notation. A local rule of radius r is written as

$$f(x_{-r}, \dots, x_r) = \sum_{i=-r}^r a_i x_i \bmod m,$$

where at least one between a_{-r} and a_r is $\neq 0$. Using this notation, the global map F of a 1-dimensional CA with radius r becomes

$$[F(c)](i) = \sum_{j=-r}^r a_j c(i+j) \bmod m, \quad c \in \mathbf{Z}_m^{\mathbf{Z}}, \quad i \in \mathbf{Z}.$$

Note that the local map f implicitly defines the neighborhood frame of the CA.

In order to study the topological properties of D -dimensional CA, we introduce a distance over the space of the configurations. Let $\Delta: \mathbf{Z}_m \times \mathbf{Z}_m \rightarrow \{0, 1\}$ defined by

$$\Delta(i, j) = \begin{cases} 0 & \text{if } i = j, \\ 1 & \text{if } i \neq j. \end{cases}$$

Given $a, b \in \mathbf{Z}_m^{\mathbf{Z}^D}$ the Tychonoff distance $d(a, b)$ is given by

$$d(a, b) = \sum_{\mathbf{v} \in \mathbf{Z}^D} \frac{\Delta(a(\mathbf{v}), b(\mathbf{v}))}{2^{\|\mathbf{v}\|_\infty}}, \quad (3)$$

where, as usual, $\|\mathbf{v}\|_\infty$ denotes the maximum of the absolute value of the components of \mathbf{v} . It is easy to verify that d is a metric on $\mathbf{Z}_m^{\mathbf{Z}^D}$ and that the metric topology induced by d coincides with the product topology induced by the discrete topology of \mathbf{Z}_m . With this topology, $\mathbf{Z}_m^{\mathbf{Z}^D}$ is a compact and totally disconnected space and F is a (uniformly) continuous map.

Throughout the paper, $F(c)$ will denote the result of the application of the map F to the configuration c , and $c(\mathbf{v})$ will denote the value assumed by c in \mathbf{v} . We recursively define $F^n(c)$ by $F^n(c) = F(F^{n-1}(c))$, where $F^0(c) = c$.

2.1. Topological properties

We now recall the definition of some dynamical properties namely, ergodicity, topological transitivity, sensitivity to initial conditions, and regularity (denseness of periodic

orbits). The above properties are widely accepted as fundamental features of chaotic behavior for general discrete time dynamical systems.

Let (X, F) be a general dynamical system, where the space of configurations X is equipped with a distance d and the map F is continuous on X according to the topology induced by d (for CA, the Tychonoff distance satisfies this property).

Definition 2.1. A dynamical system (X, F) is topologically transitive if and only if for all nonempty open subsets U and V of X there exists a natural number n such that $F^n(U) \cap V \neq \emptyset$.

Intuitively, a topologically transitive map has points which eventually move under iteration from one arbitrarily small neighborhood to any other. As a consequence, the dynamical system cannot be decomposed into two disjoint open sets which are invariant under the map.

Definition 2.2. A dynamical system (X, F) is sensitive to initial conditions if and only if there exists a $\delta > 0$ such that for any $x \in X$ and for any neighborhood $N(x)$ of x , there is a point $y \in N(x)$ and a natural number n , such that $d(F^n(x), F^n(y)) > \delta$. δ is called the sensitivity constant.

Intuitively, a map is sensitive to initial conditions, or simply sensitive, if there exist points arbitrarily close to x which eventually separate from x by at least δ under iteration of F . We emphasize that not all points near x need eventually separate from x under iteration, but there must be at least one such point in every neighborhood of x .

If a map possesses sensitive dependence on initial conditions, then, for all practical purposes, the dynamics of the map defies numerical approximation. Small errors in computation which are introduced by round-off may become magnified upon iteration. The results of numerical computation of an orbit, no matter how accurate, may be completely different from the real orbit.

In [4] it has been proven that, for CA, topological transitivity implies sensitivity. Thus, for CA, the notion of transitivity becomes central to chaos theory.

Definition 2.3. Let $P(F) = \{x \in X \mid \exists n \in \mathbb{N} : F^n(x) = x\}$ be the set of the periodic points of F . A dynamical system (X, F) has dense periodic orbits if and only if $P(F)$ is a dense subset of X , i.e., for any $x \in X$ and $\varepsilon > 0$, there exists $y \in P(F)$ such that $d(x, y) < \varepsilon$.

Denseness of periodic orbits is often referred to as the *element of regularity* a chaotic dynamical system must exhibit. For this reason we say that a dynamical system is *regular* if it has dense periodic orbits. The popular book by Devaney [8] isolates three components as being the essential features of chaos: transitivity,

sensitivity to initial conditions and regularity. Finally, we recall the definition of ergodic map.

Definition 2.4. Let (X, \mathcal{F}, μ) be a probability space. Let $F : X \rightarrow X$, be a measurable map which preserves μ , i.e., for every subset $E \in \mathcal{F}$ we have $\mu(E) = \mu(F^{-1}(E))$. Then F is ergodic with respect to μ if and only if for every $E \in \mathcal{F}$

$$(E = F^{-1}(E)) \Rightarrow (\mu(E) = 0 \text{ or } \mu(E) = 1).$$

In order to apply ergodic theory to CA we need to define \mathcal{F} , i.e., the collection of measurable subsets of $\mathbf{Z}_m^{\mathbf{Z}^D}$ and a probability measure $\mu : \mathcal{F} \rightarrow [0, 1]$. We will use the normalized *Haar* measure μ_H defined over the σ -algebra of *cylinders* which is, to our knowledge, one of the most widely used probabilistic setting in CA theory. Formally, let $(\mathbf{Z}_m^{\mathbf{Z}^D}, F)$ be a CA:

- μ_H is defined as the product measure induced by the uniform probability distribution over A and
- a D -dimensional *cylinder* $\langle (v_1, a_1), \dots, (v_l, a_l) \rangle$ is a particular subset of $\mathbf{Z}_m^{\mathbf{Z}^D}$ defined as

$$\langle (v_1, a_1), \dots, (v_l, a_l) \rangle = \left\{ x \in \mathbf{Z}_m^{\mathbf{Z}^D} : x(v_i) = a_i, i = 1, \dots, l \right\},$$

where $a_i \in \mathbf{Z}$ and $v_i \in \mathbf{Z}^D$.

One can easily verify that

$$\mu_H(\langle (v_1, a_1), \dots, (v_l, a_l) \rangle) = \frac{1}{m^l},$$

where m is the cardinality of A . Note that cylinders form a basis of closed and open (clopen from now) subsets of $\mathbf{Z}_m^{\mathbf{Z}^D}$ according to the metric topology induced by the Tychonoff distance. Since in the rest of this paper we will only use Haar measure, then we will write μ instead of μ_H .

3. Statement of new results

In this section we state the main results of this paper. The same results are summarized in the diagram of Fig. 1.

Our first result shows that transitive CA (hence also ergodic CA in view of Theorem 4.3) are surjective.

Theorem 3.1. *Topologically transitive CA with respect to the metric topology induced by the Tychonoff distance are surjective.*

The following theorem shows that for linear CA there exists a simple characterization of topologically transitive and of ergodic maps based on the coefficients of the local rule.

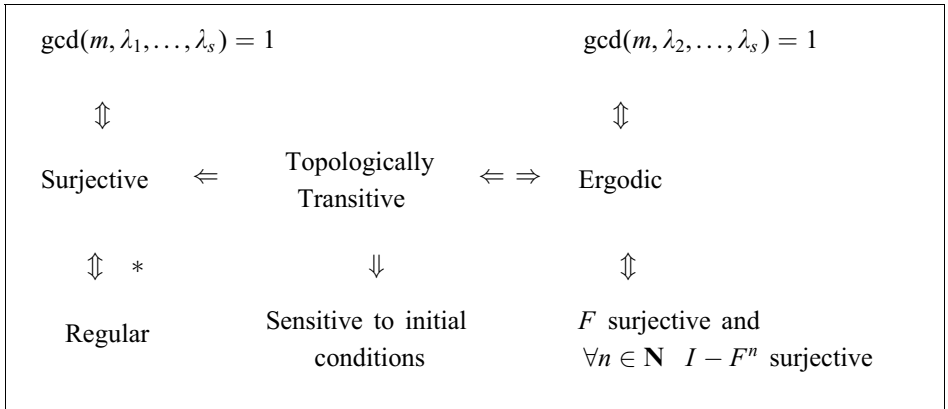


Fig. 1. Relations among properties of global transition maps associated to D -dimensional linear CA over \mathbf{Z}_m . Implication marked with * has been proved only in the 1-dimensional case. The following relations hold for general CA: transitivity \Rightarrow sensitivity, transitivity \Rightarrow surjectivity, and ergodicity \Rightarrow transitivity.

Theorem 3.2. *Let F denote the global transition map of the additive D -dimensional CA over \mathbf{Z}_m defined by*

$$[F(c)](\mathbf{v}) = \sum_{i=1}^s \lambda_i c(\mathbf{v} + \mathbf{u}_i). \quad (4)$$

Assume $\mathbf{u}_1 = \mathbf{0}$, that is, λ_1 is the coefficient associated to the null displacement. The following statements are equivalent:

- (a) *F is ergodic,*
- (b) *F is topologically transitive,*
- (c) *$\gcd(m, \lambda_2, \lambda_3, \dots, \lambda_s) = 1$, where \gcd denotes the greatest common divisor.*

Note that for a linear 1-dimensional CA with local rule $f(x_{-r}, \dots, x_r) = \sum_{i=-r}^r a_i x_i \bmod m$, the above theorem states that

$$F \text{ is ergodic} \Leftrightarrow F \text{ is top. transitive} \Leftrightarrow \gcd(m, a_{-r}, \dots, a_{-1}, a_1, \dots, a_r) = 1.$$

Theorem 3.2 generalizes a result given in [16] where the authors prove that 1-dimensional linear CA over \mathbf{Z}_2 based on a local rule different from the identity map are ergodic. In addition, our result makes it possible to give a complete answer the following question raised in [17]: *Are all surjective 1-dimensional CA defined over \mathbf{Z}_2 (with the exception of the identity and the inversion map) ergodic?* For the class of linear CA over \mathbf{Z}_m , Theorem 3.2 tells us that this is true if and only if m is prime. For general CA, the statement is false even for m prime. In fact, in [1] the authors show that there exists a non trivial 1-dimensional CA over \mathbf{Z}_2 , such that the global map F is invertible and $F = F^{-1}$. Clearly, F must be surjective but, since F^2 is the identity map, it is not ergodic.

Our final result shows that 1-dimensional linear CA over \mathbf{Z}_m are regular if and only if they are surjective.

Theorem 3.3. *Let F be a 1-dimensional additive CA over \mathbf{Z}_m . The set $P(F)$ of periodic points of F is a dense subset of $\mathbf{Z}_m^{\mathbf{Z}}$ (F is regular) iff F is surjective.*

This result is a generalization of Theorem 4 of [9] where it has been proved that 1-dimensional linear CA over \mathbf{Z}_p with p prime have dense periodic points.

4. Proof of the main theorems

We now prove the results stated in Section 3. In our proofs we make use of the following known facts about CA.

Theorem 4.1 (Ito, Osato and Nasu [13]). *Let $f: \mathbf{Z}_m^s \rightarrow \mathbf{Z}_m$ be the linear map defined by*

$$f(x_1, \dots, x_s) = \sum_{i=1}^s \lambda_i x_i \bmod m.$$

The CA based on the local rule f is surjective if and only if $\gcd(m, \lambda_1, \dots, \lambda_s) = 1$.

Theorem 4.2 (Shirvani and Rogers [16]). *Let G be a compact abelian group with normalized Haar measure μ , and let θ be a continuous surjective endomorphism of G . Then, θ is ergodic iff for every $n \geq 1$, $\theta^n - I$ is surjective.*

Theorem 4.3 (Willson [18]). *Ergodic CA with respect to the normalized Haar measure are topologically transitive with respect to the metric topology induced by the Tychonoff distance.*

4.1. Ergodicity and topological transitivity

This section contains the proof of Theorems 3.1 and 3.2.

Proof of Theorem 3.1. Assume by contradiction that F is not surjective, that is, there exists $y \in X$ such that, for all $x \in X$, $F(x) \neq y$. Since F is topologically transitive, we can find a sequence x_n such that $d(F(x_n), y) < 1/n$. Since (X, d) is a compact metric space, we can find a subsequence x_{n_i} which converges to $z \in X$. We have

$$d(F(z), y) \leq d(F(z), F(x_{n_i})) + d(F(x_{n_i}), y).$$

Since the right-hand term goes to zero as $n_i \rightarrow \infty$ we have $F(z) = y$ which contradicts our hypothesis. Hence F must be surjective as claimed. \square

For the study of additive CA, we make use of the *formal power series* (fps) representation of the configuration space $\mathbf{Z}_m^{\mathbf{Z}^D}$ (see [13, Section 3] for details). For $D = 1$, to each configuration $c \in \mathbf{Z}_m^{\mathbf{Z}}$ we associate the fps

$$P_c(X) = \sum_{i \in \mathbf{Z}} c(i)X^i.$$

The advantage of this representation is that the computation of an additive map is equivalent to power series multiplication. Let $F : \mathbf{Z}_m^{\mathbf{Z}} \rightarrow \mathbf{Z}_m^{\mathbf{Z}}$ be an additive map with local rule $f(x_{-r}, \dots, x_r) = \sum_{i=-r}^r a_i x_i$. We associate to F the finite fps $A_f(X) = \sum_{i=-r}^r a_i X^{-i}$. Then, for any $c \in \mathbf{Z}_m^{\mathbf{Z}}$ we have

$$P_{F(c)}(X) = P_c(X)A_f(X).$$

Note that each coefficient of $P_{F(c)}(X)$ is well defined since $A_f(X)$ has only finitely many nonzero coefficients. Note also that the finite fps associated to F^n is $[A_f(X)]^n$.

More in general, to each configuration $c \in \mathbf{Z}_m^{\mathbf{Z}^D}$ we associate the formal power series

$$P_c(X_1, \dots, X_D) = \sum_{(i_1, \dots, i_D) \in \mathbf{Z}^D} c(i_1, \dots, i_D) X_1^{i_1} \cdots X_D^{i_D}.$$

The computation of an additive map F over \mathbf{Z}_m is equivalent to the multiplication by a finite fps $A(X_1, \dots, X_D)$ which is obtained by the local rule f and the neighbourhood frame $\mathbf{u}_1, \dots, \mathbf{u}_s$ as follows. The finite fps associated to the map F defined by (4) is

$$A(X_1, \dots, X_D) = \sum_{i=1}^s \lambda_i X_1^{-\mathbf{u}_i(1)} \cdots X_D^{-\mathbf{u}_i(D)}, \quad (5)$$

where $\mathbf{u}_i(j)$ denotes the j -th component of vector \mathbf{u}_i . As for 1-dimensional CA, the finite fps associated to F^n is $[A(X_1, \dots, X_D)]^n$. Since finite fps's multiply with the same rule as polynomials, the coefficient of the monomial $X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_D^{\alpha_D}$ in $[A(X_1, \dots, X_D)]^n$ is given by

$$A_{\alpha_1 \alpha_2 \dots \alpha_D} = \sum_{1 \leq i_1, \dots, i_n \leq D} \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_n}, \quad (6)$$

where the summation is restricted to the set of indices i_1, \dots, i_n such that, for $j = 1, \dots, D$,

$$\mathbf{u}_{i_1}(j) + \mathbf{u}_{i_2}(j) + \cdots + \mathbf{u}_{i_n}(j) = -\alpha_j.$$

In other words, the coefficient of the monomial $X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_D^{\alpha_D}$ is equal to the sum of all n -term products $\lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_n}$ such that the sum of the displacements $\mathbf{u}_{i_1} + \cdots + \mathbf{u}_{i_n}$ is equal to $-(\alpha_1, \alpha_2, \dots, \alpha_D)$. The minus sign is due to the presence of the minuses at the exponents of the X_i 's in (5).

Proof of Theorem 3.2. We first prove (a) is equivalent to (c). Let F be the global transition map of the additive D -dimensional CA given by (4). We assume that $\mathbf{u}_1 = 0$,

and that $\mathbf{u}_i \neq \mathbf{u}_j$ for $i \neq j$. Let

$$A(X_1, \dots, X_D) = \lambda_1 + \sum_{i=2}^s \lambda_i X_1^{-\mathbf{u}_i(1)} \dots X_D^{-\mathbf{u}_i(D)}$$

denote the finite fps associated to F . We first prove that $\gcd(m, \lambda_2, \dots, \lambda_s) = 1$ implies that F is ergodic. By Theorem 4.1 we know that F is surjective. We prove our thesis by showing that $F^n - I$ is surjective for all $n \geq 1$ and using Theorem 4.2.

Let \mathcal{P} denote the set of all primes factor of m . A fundamental observation is that $F^n - I$ is surjective iff for each $p \in \mathcal{P}$, p does not divide all coefficients of the finite fps $[A(X_1, \dots, X_D)]^n - 1$. In other words

$$F^n - I \text{ surjective} \Leftrightarrow \forall p \in \mathcal{P} [A(X_1, \dots, X_D)]^n \not\equiv 1 \pmod{p}. \quad (7)$$

Let $\|\mathbf{v}\|_2$ denote the euclidean norm of vector \mathbf{v} . For any $p \in \mathcal{P}$, we consider a displacement \mathbf{u}_j such that $\|\mathbf{u}_j\|_2 \geq \|\mathbf{u}_i\|_2$ for all i such that $\lambda_i \not\equiv 0 \pmod{p}$ (note that \mathbf{u}_j is not necessarily unique). Since $\gcd(m, \lambda_2, \dots, \lambda_s) = 1$, p cannot divide all the coefficients λ_i 's for $i \geq 2$. Hence, $j \geq 2$ and $\|\mathbf{u}_j\|_2 \geq 1$. We prove that $[A(X_1, \dots, X_D)]^n \not\equiv 1 \pmod{p}$ by showing that it contains the monomial $X_1^{-n\mathbf{u}_j(1)} \dots X_D^{-n\mathbf{u}_j(D)}$ whose coefficient $C_{n,j}$ is nonzero modulo p .

In view of (6), $C_{n,j}$ is given by the sum of all n -term products $\lambda_{i_1} \lambda_{i_2} \dots \lambda_{i_n}$ such that the sum of the displacements $\mathbf{u}_{i_1} + \dots + \mathbf{u}_{i_n}$ is equal to $n\mathbf{u}_j$. Since $\lambda_i \not\equiv 0 \pmod{p}$ implies $\|\mathbf{u}_i\|_2 \leq \|\mathbf{u}_j\|_2$, the term $\lambda_{i_1} \lambda_{i_2} \dots \lambda_{i_n}$ gives a nonzero (modulo p) contribution to $C_{n,j}$ if and only if

$$\mathbf{u}_{i_1} + \dots + \mathbf{u}_{i_n} = n\mathbf{u}_j \quad \text{and} \quad \|\mathbf{u}_{i_k}\|_2 \leq \|\mathbf{u}_j\|_2 \quad \text{for } k = 1, 2, \dots, n. \quad (8)$$

Since for the euclidean norm we have $\|\mathbf{v} + \mathbf{u}\|_2 < \|\mathbf{v}\|_2 + \|\mathbf{u}\|_2$ unless \mathbf{u} and \mathbf{v} have the same direction, (8) holds iff $\mathbf{u}_{i_1} = \mathbf{u}_{i_2} = \dots = \mathbf{u}_{i_n} = \mathbf{u}_j$. Hence, $C_{n,j} \equiv \lambda_j^n \pmod{p}$ and $C_{n,j} \not\equiv 0 \pmod{p}$ as claimed. Since the same reasoning holds for any $p \in \mathcal{P}$ we have that $F^n - I$ is surjective for all $n \geq 1$ and, by Theorem 4.2, F is ergodic.

We now prove that $\gcd(m, \lambda_2, \dots, \lambda_s) = 1$ is a necessary condition for ergodicity. Assume by contradiction that F is ergodic and $\gcd(m, \lambda_2, \dots, \lambda_s) > 1$. Since ergodicity implies surjectivity (Theorems 4.3 and 3.1), by Theorem 4.1 we know that $\gcd(m, \lambda_1, \dots, \lambda_s) = 1$. Hence, there exists a prime p such that

$$p|m, \quad p \nmid \lambda_1, \quad \text{and} \quad p|\lambda_i \text{ for } i = 2, \dots, s.$$

By (7), to prove that F cannot be ergodic it suffices to show that there exists n such that $[A(X_1, \dots, X_D)]^n \equiv 1 \pmod{p}$. By choosing $n = p - 1$ we have

$$[A(X_1, \dots, X_D)]^{p-1} \equiv (\lambda_1)^{p-1} \equiv 1 \pmod{p}$$

where the last equality follows from the Fermat theorem which states that $x^{p-1} \equiv 1 \pmod{p}$ when p is prime and $x \not\equiv 0 \pmod{p}$.

Since ergodic CA are topologically transitive (Theorem 4.3) to complete the proof it suffices to show that (b) implies (c). Assume by contradiction that F is topologically

transitive and $\gcd(m, \lambda_2, \dots, \lambda_s) = d > 1$. Since F is surjective (Theorem 3.1) by Theorem 4.1 we must have $\gcd(m, \lambda_1, \dots, \lambda_s) = 1$. By (2), for any $c \in \mathbf{Z}_m^{\mathbf{Z}^D}$ and $\mathbf{v} \in \mathbf{Z}^D$ we have

$$[F(c)](\mathbf{v}) = \lambda_1 c(\mathbf{v}) + \sum_{i=1}^s \lambda_i c(\mathbf{v} + \mathbf{u}_i) \bmod m.$$

By hypothesis, d divides every λ_i for $i > 2$, while $\gcd(\lambda_1, d) = 1$. Hence, for the properties of the gcd, we have

$$\gcd([F(c)](\mathbf{v}), d) = \gcd(\lambda_1 c(\mathbf{v}) + dH, d) = \gcd(\lambda_1 c(\mathbf{v}), d) = \gcd(c(\mathbf{v}), d). \quad (9)$$

In other words, $\gcd(c(\mathbf{v}), d)$ is invariant under iteration of F . Consider the cylinders $C_0 = \langle (\mathbf{0}, 0) \rangle$ and $C_1 = \langle (\mathbf{0}, 1) \rangle$. By (9), for any $c \in C_0$ and $n > 0$ we have $\gcd([F^n(c)](\mathbf{0}), d) = \gcd(c(\mathbf{0}), d) = d$. Hence, $F^n(C_0) \cap C_1 = \emptyset$ and F cannot be topologically transitive.

This completes the proof. \square

4.2. Regularity

In this section we prove that for 1-dimensional linear CA over \mathbf{Z}_m regularity is equivalent to surjectivity (Theorem 3.3). In view of Theorem 4.1 we conclude that a 1-dimensional linear CA over \mathbf{Z}_m is regular if and only if the greatest common divisor of all the coefficients of the local rule associated to the CA and m is equal to 1. We conjecture that the above characterization of regularity still holds in the D -dimensional case. Unfortunately, the proof technique we use in the 1-dimensional case cannot be extended to the D -dimensional case.

To prove Theorem 3.3 we need some preliminary definitions and lemmas. Let $\sigma : \mathbf{Z}_m^{\mathbf{Z}} \rightarrow \mathbf{Z}_m^{\mathbf{Z}}$, defined by

$$\forall x \in \mathbf{Z}_m^{\mathbf{Z}} \quad \forall i \in \mathbf{Z} \quad [\sigma(x)](i) = x(i+1),$$

denote the 1-dimensional *shift* CA. We say that a configuration $x \in \mathbf{Z}_m^{\mathbf{Z}}$ is spatially periodic if and only if there exists $s \in \mathbf{N}$ such that $\sigma^s(x) = x$. We have the following result (which is a generalization of Theorem 2.3, p. 279 of [5]).

Lemma 4.4. *Let F be a surjective 1-dimensional linear CA over \mathbf{Z}_m . Every predecessor of a spatially periodic configuration is spatially periodic*

Proof. Let $x, y \in \mathbf{Z}_m^{\mathbf{Z}}$ be such that $F(x) = y$ and $\sigma^s(y) = y$ for some $s \in \mathbf{N}$. For every $i \in \mathbf{Z}$ we have

$$F(\sigma^{is}(x)) = \sigma^{is}(F(x)) = \sigma^{is}(y) = y.$$

Assume that x is not spatially periodic. Then there exist infinitely many predecessors of y according to F namely, $\sigma^{is}(x)$, $i \in \mathbf{Z}$. Since every 1-dimensional surjective CA

Proof. Let a_{-r}, \dots, a_r be the coefficients of the local rule f associated to F . Since F is a Right CA, we have $a_0 = 0$. Let b_{-r}, \dots, b_r be the coefficients of the local rule g associated to G . Since $G = I - F$, we have $b_0 = 1$, hence $\gcd(m, b_{-r}, \dots, b_0, \dots, b_r) = 1$ and by Theorem 4.1 G is surjective. \square

We now prove that for surjective right (left) linear CA $(\mathbf{Z}_m^{\mathbf{Z}}, F)$ every periodic configuration for F is periodic also for σ , i.e., is spatially periodic.

Theorem 4.7. *Let F be a surjective Right [Left] linear CA over \mathbf{Z}_m . Then for every $x \in \mathbf{Z}_m^{\mathbf{Z}}$ we have*

$$(\exists t \in \mathbf{N} : F^t(x) = x) \Rightarrow (\exists s \in \mathbf{N} : \sigma^s(x) = x).$$

Proof. If x is periodic for F , i.e., $F^t(x) = x$, then x is a predecessor of the all-zero configuration $\mathbf{0}$ according to $G = I - F^n$. From Lemma 4.6 we have that $G = I - F^n$ is surjective for every $n \in \mathbf{N}$. From Lemma 4.4 we conclude that x is spatially periodic. \square

We have the following corollary.

Corollary 4.8. *Let F be a surjective 1-dimensional linear CA over \mathbf{Z}_m . Let $n \in \mathbf{Z}$ be such that $G = \sigma^n F$ is a Right [Left] CA. Every periodic configuration for G is periodic also for F , i.e.,*

$$(\exists t \in \mathbf{N} : G^t(x) = x) \Rightarrow (\exists t' \in \mathbf{N} : F^{t'}(x) = x).$$

Proof. From Theorem 4.7 we have that there exists $s \in \mathbf{N}$ such that $\sigma^s(x) = x$. We have

$$\begin{aligned} x &= G^t(x) = G^{ts}(x) = (\sigma^n F)^{ts}(x) \\ &= \sigma^{nts} F^{ts}(x) = F^{ts} \sigma^{nts}(x) = F^{ts}(x). \end{aligned} \quad \square$$

We recall the definition of *cross section* which will be used in the proof of Lemma 4.11.

Definition 4.9 (Hedlund [11]). Let X and Y be topological spaces and let $f : X \rightarrow Y$, be continuous. Then $g : Y \rightarrow X$, is a *cross-section* of f provided g is continuous and

$$\forall y \in Y : f(g(y)) = y.$$

Hedlund in [11] proved the following result.

Theorem 4.10 (Hedlund [11]). *Let F be a 1-dimensional CA. The following statements are equivalent.*

- (i) *Every configuration x has the same number of predecessors.*
- (ii) *F has a cross section.*

$$\begin{array}{c}
 x = \dots 000000 \text{*****} \dots \\
 \downarrow F \\
 e_0 = \dots 00000000000010000000000 \dots
 \end{array}$$

Fig. 3. Let F be a surjective 1-dimensional linear CA over \mathbf{Z}_m . There exists $k \in \mathbf{Z}$ such that the configuration e_0 has a predecessor x with $x(i) = 0$ for every $i < k$.

Let $e_i \in \mathbf{Z}_m^{\mathbf{Z}}$ be defined as follows.

$$e_i(j) = \begin{cases} 1 & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

We are now ready to prove the following lemma (see Fig. 3).

Lemma 4.11. *Let F be a surjective 1-dimensional linear CA over \mathbf{Z}_m . Then there exist $x \in \mathbf{Z}_m^{\mathbf{Z}}$ and $k \in \mathbf{Z}$ such that*

$$F(x) = e_0 \quad \text{and} \quad \forall i < k : x(i) = 0.$$

Proof. Since F is a surjective linear CA, every configuration of F has the same (finite) number of predecessors. From Theorem 4.10 we have that F has a cross section g . Since g is continuous, we have

$$\lim_{n \rightarrow +\infty} d(e_n, \mathbf{0}) = 0 \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} d(g(e_n), g(\mathbf{0})) = 0. \quad (10)$$

Let $\alpha_n = g(e_n)$ and $\beta = g(\mathbf{0})$. By Equation 9 we have

$$\forall k \in \mathbf{N} \quad \exists n > k : \quad \alpha_n(i) = \beta(i), \quad -k \leq i \leq k. \quad (11)$$

Let

$$M_n = \max\{k \in \mathbf{N} : \beta(i) = \alpha_n(i), \quad -k \leq i \leq k\}$$

and

$$\gamma_n(i) = \begin{cases} \beta(i) & \text{if } i \leq -M_n, \\ \alpha_n(i) & \text{otherwise.} \end{cases}$$

It is easy to verify that there exists \tilde{n} large enough such that $F(\gamma_{\tilde{n}}) = e_{\tilde{n}}$ and

$$\forall i \leq -M_{\tilde{n}} : \quad \gamma_{\tilde{n}}(i) = \beta(i). \quad (12)$$

Let $\alpha = \gamma_{\tilde{n}}$ and $x = (m-1)\beta + \alpha$. We have

$$\begin{aligned}
 F(x) &= F((m-1)\beta + \alpha) \\
 &= (m-1)F(\beta) + F(\alpha) \\
 &= (m-1)F(g(\mathbf{0})) + F(\gamma_{\tilde{n}})
 \end{aligned}$$

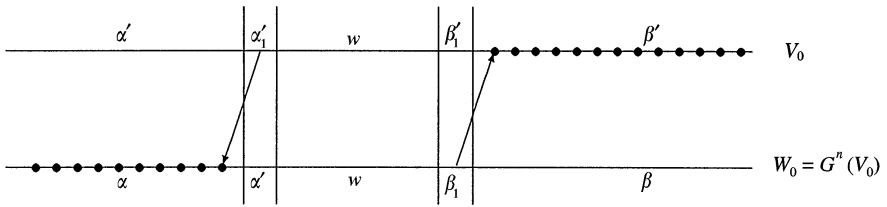


Fig. 4. α'_1 and β_1 can be arbitrarily modified without influencing α_1 and β'_1 , respectively.

$$= (m-1)\mathbf{0} + e_{\tilde{n}}$$

$$= e_{\tilde{n}}.$$

Since F is shift invariant, we have

$$F\sigma^{\tilde{n}}(x) = \sigma^{\tilde{n}}F(x) = \sigma^{\tilde{n}}e_{\tilde{n}} = e_0.$$

From Eq. (11) we have

$$\forall i < -M_{\tilde{n}} - \tilde{n} : x(i) = (m-1)\beta(i) + \alpha(i) = 0$$

as claimed. \square

We are now ready to prove Theorem 3.3.

Proof of Theorem 3.3. Assume that F is surjective. Let $h \in \mathbb{Z}$ be such that

- (i) $G = \sigma^h F$ is a Right CA, and
- (ii) there exists $x \in \mathbb{Z}_m^{\mathbb{Z}}$ such that

$$G(x) = e_0 \quad \text{and} \quad \forall i < 1 : x(i) = 0.$$

An integer h which satisfies properties (i) and (ii) does exist in view of Lemma 4.11.

Let $w \in \mathbb{Z}_m^{2k+1}$ be any finite configuration of length $2k+1$, where k is an arbitrarily chosen positive integer. Since G is a surjective Right CA, Theorem 3.3 implies that G is topologically transitive. Hence, there exist $n \in \mathbb{N}$ and $V_0, W_0 \in \mathbb{Z}_m^{\mathbb{Z}}$ such that

$$V_0 = \cdots \alpha_2 \alpha_1 w_{-k} \cdots w_0 \cdots w_k \beta_1 \beta_2 \cdots$$

$$W_0 = \cdots \alpha'_2 \alpha'_1 w_{-k} \cdots w_0 \cdots w_k \beta'_1 \beta'_2 \cdots$$

and

$$G^n(V_0) = W_0.$$

Note that w is centered at the origin of the lattice, i.e., $V_0(0) = W_0(0) = w_0$.

From property (ii) and from the linearity of G we have that there exists a configuration $y \in \mathbb{Z}_m^{\mathbb{Z}}$ such that

$$G^n(y) = e_0 \quad \text{and} \quad \forall i < 1 : y(i) = 0.$$

Consider now the following configuration

$$Q = V_0 + (\beta'_1 - \beta_1)\sigma^{-k-1}(y).$$

Since $G^n(y) = e_0$ we have

$$G^n(\sigma^{-k-1}(y)) = \sigma^{-k-1}(G^n(y)) = \sigma^{-k-1}(e_0) = e_{k+1}$$

and then

$$\begin{aligned} G^n(Q) &= G^n(V_0 + (\beta'_1 - \beta_1)\sigma^{-k-1}(y)) \\ &= G^n(V_0) + (\beta'_1 - \beta_1)G^n(\sigma^{-k-1}(y)) \\ &= W_0 + (\beta'_1 - \beta_1)e_{k+1} \\ &= \cdots \alpha_2 \alpha_1 w_{-k} \cdots w_0 \cdots w_k \beta'_1 \beta_2 \cdots \end{aligned}$$

Let

$$V_1 = Q + (\alpha'_1 - \alpha_1)e_{-k-1}.$$

We have

$$\begin{aligned} G^n(V_1) &= G^n(Q + (\alpha'_1 - \alpha_1)e_{-k-1}) \\ &= G^n(Q) + (\alpha'_1 - \alpha_1)G^n(e_{-k-1}) \\ &= \cdots \gamma_2 \gamma_1 \alpha'_1 w_{-k} \cdots w_0 \cdots w_k \beta'_1 \beta_2 \cdots, \end{aligned}$$

where $\gamma(i) \in \mathbf{Z}_m$, $i \in \mathbf{N}$, and

$$V_1(j) = W_1(j), \quad j = -k-1, \dots, k+1.$$

By repeating the above procedure we are able to construct a sequence of pairs of configurations (V_i, W_i) such that $G^n(V_i) = W_i$ and $V_i(j) = W_i(j)$ for $j = -i-k, \dots, k+i$ and $i = 1, 2, \dots$. Since $\mathbf{Z}_m^{\mathbf{Z}}$ is a complete space we have

$$\lim_{i \rightarrow \infty} W_i = \lim_{i \rightarrow \infty} V_i = W \quad \text{and} \quad G^n(W) = W.$$

Since w can be arbitrarily chosen, we conclude that G has dense periodic orbits. Finally, by Corollary 4.8 we conclude that F has dense periodic orbits.

It remains to prove that if F has dense periodic points then it is surjective. Assume by contradiction that F is not surjective, that is, there exists $y \in \mathbf{Z}_m^{\mathbf{Z}}$ such that, for all $x \in \mathbf{Z}_m^{\mathbf{Z}}$, $F(x) \neq y$. Since F has dense periodic points, we can find a sequence of periodic configurations $x_n \in \mathbf{Z}_m^{\mathbf{Z}}$ of period p_n such that

$$d(x_n, y) = d(F^{p_n}(x_n), y) < 1/n.$$

Consider now the sequence $x'_n = F^{p_n-1}(x_n)$. Since $\mathbf{Z}_m^{\mathbf{Z}}$ is a compact metric space, we can extract from x'_n a subsequence x''_{n_i} which converges to $z \in \mathbf{Z}_m^{\mathbf{Z}}$. We have

$$d(F(z), y) \leq d(F(z), F(x''_{n_i})) + d(F(x''_{n_i}), y).$$

Since the right-hand term goes to zero as $n_i \rightarrow \infty$ we have $F(z) = y$ which contradicts our hypothesis. Hence, F must be surjective as claimed. \square

5. Conclusions and further works

We have solved the problem of deciding when a given D -dimensional additive CA over \mathbf{Z}_m is ergodic. We have shown that this problem is computationally equivalent to compute the greatest common divisor of some of the coefficients of the local rule on which the CA is based. We have also proved that an additive CA is ergodic if and only if it is topologically transitive. This last result establishes a connection between ergodic theory and topological chaos for CA. Finally, we have proved that in the case of 1-dimensional additive CA over \mathbf{Z}_m , regularity is equivalent to surjectivity. It remains open, among the others, the two following problems:

1. regularity is equivalent to surjectivity also for D -dimensional additive CA over \mathbf{Z}_m , or more in general for surjective CA?
2. topologically transitive CA are ergodic?

References

- [1] S. Amoroso, Y.N. Patt, Decision procedures for surjectivity and injectivity of parallel maps for tessellation structures, *J. Comput. System Sci.* 6 (1972) 448–464.
- [2] H. Aso, N. Honda, Dynamical characteristics of linear cellular automata, *J. Comput. System Sci.* 30 (1985) 291–317.
- [3] G. Cattaneo, E. Formenti, G. Manzini, L. Margara, On ergodic linear cellular automata over \mathbf{Z}_m , in: 14th Annu. Symp. on Theoretical Aspects of Computer Science, *Lecture Notes in Computer Science*, vol. 1200, Springer, 1997, Berlin, pp. 427–438.
- [4] B. Codenotti, L. Margara, Transitive cellular automata are sensitive, *Amer. Math. Monthly* 103 (1996) 58–62.
- [5] R. Cordovil, R. Dilao, A. Noronha da Costa, Periodic orbits for additive cellular automata, *Discrete Comput. Geom.* 1 (3) (1986) 277–288.
- [6] K. Culik, J. Pachl, S. Yu, On the limit sets of cellular automata, *SIAM J. Comput.* 18 (1989) 831–842.
- [7] K. Culik, S. Yu, Undecidability of CA classification schemes, *Complex Systems* 2 (1988) 177–190.
- [8] R.L. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd ed., Addison-Wesley, Reading, MA, 1989.
- [9] P. Favati, G. Lotti, L. Margara, One dimensional additive cellular automata are chaotic according to Devaney's definition of chaos, *Theoret. Comput. Sci.* 174 (1997) 157–170.
- [10] P. Guan, Y. He, Exact results for deterministic cellular automata with additive rules, *J. Stat. Phys.* 43 (1986) 463–478.
- [11] G.A. Hedlund, Endomorphisms and automorphisms of the shift dynamical system, *Math. Systems Theory* 3 (1969) 320–375.
- [12] M. Hurley, Ergodic aspects of cellular automata, *Ergodic theory and dynamical systems* 10 (1990) 671–685.

- [13] M. Ito, N. Osato, M. Nasu, Linear cellular automata over Z_m , *J. Comput. System Sci.* 27 (1983) 125–140.
- [14] J. Kari, Rice's theorem for the limit set of cellular automata, *Theoret. Comput. Sci.* 127 (2) (1994) 229–254.
- [15] T. Sato, Ergodicity of linear cellular automata over Z_m , *Inform. Process. Lett.* 61 (3) (1997) 169–172.
- [16] M. Shirvani, T.D Rogers, Ergodic endomorphisms of compact abelian groups, *Commun. Math. Phys.* 118 (1988) 401–410.
- [17] M. Shirvani, T.D Rogers, On ergodic one-dimensional cellular automata, *Commun. Math. Phys.* 136 (3) (1991) 599–605.
- [18] S.J. Willson, On the ergodic theory of cellular automata, *Math. Systems Theory* 9 (1975) 132–141.